

MAGDALENA SZPUNAR

Internet in the process of secondary data collecting

Internet w procesie gromadzenia danych o charakterze wtórnym

KEY WORDS

secondary data, data collecting
process, advantages and disadvantages
of secondary data, secondary data quality analysis

SŁOWA KLUCZOWE

dane wtórne, proces gromadzenia danych, zalety
i wady danych wtórnych, analiza jakości danych
wtórnych

ABSTRACT

Secondary data is a specific type of data as it is information from already existing sources which does not require researchers to conduct the process of initial data gathering and research. The Internet is a typical example of such a source. The article discusses the advantages and disadvantages of such data with a focus on secondary data that can be found on the Web. Additionally, the author also presents ways of assessment of such data reliability and credibility.

STRESZCZENIE

Dane wtórne stanowią specyficzny typ danych, pozwalają bowiem pozyskiwać je ze źródeł już istniejących, nie wymagając od badacza przeprowadzania badań. Coraz częściej źródłem tego typu danych staje się Internet. Artykuł rozważa zalety i wady danych wtórnych, ze szczególnym wskazaniem na dane zastane, które możemy znaleźć w Internecie. Autorka są również przedstawia również sposoby ułatwiające ocenę wiarygodności i rzetelności danych zastanych, które znajdują się w globalnej Sieci.

Wprowadzenie

Analitik, próbując udzielić odpowiedzi na pytanie określone w problematyce badawczej, może podążać dwiema drogami. Może zrealizować własne badania, uzyskując tzw. dane pierwotne, może również skorzystać z danych zebranych przez inne osoby, korzystając w ten sposób z tzw. danych wtórnych czy, inaczej mówiąc, zastanych. Jak podaje Earl Babbie, analiza wtórna to taka forma badań, gdzie dane zebrane i przetworzone przez jednego badacza są ponownie analizowane przez innego¹. Jak podkreśla Krystyna Mazurek-Łopacińska, wybór do późniejszej analizy danych zgromadzonych, przechowywanych i przetwarzanych przez inne osoby, czyli danych z tzw. drugiej ręki, wymaga jednak umiejętności ich selekcji i redukcji². Gilbert Churchill twierdzi, że dopiero wtedy, kiedy dane wtórne (zastane) zostaną wyczerpane, gdy badacz nie odnosi z nich większych korzyści, powinien on rozpocząć proces tworzenia danych pierwotnych. Autor ten uważa, że badania sondażowe winny być podejmowane wyłącznie wówczas, gdy zostaną wykorzystane wszelkie inne możliwości; podobnie jak chirurg, który decyduje się na wykonanie zabiegu jedynie w ostateczności³. Choć trudno stawiać tutaj kategorię sądów, każdy bowiem problem badawczy wymaga innej metody czy techniki analizy, to jednak zawsze warto dokładnie przemyśleć, czy nie istnieją dane, które mogą pomóc w udzieleniu odpowiedzi na nasze pytanie badawcze.

¹ E. Babbie, *Badania społeczne w praktyce*, Warszawa 2004, s. 303.

² K. Mazurek-Łopacińska, *Gromadzenie danych wtórnych*, [w:] *Badania marketingowe. Teoria i praktyka*, pod red. K. Mazurek-Łopacińskiej, Warszawa 2005, s. 58.

³ G.A. Churchill, *Badania marketingowe. Podstawy metodologiczne*, Warszawa 2002, s. 250–252.

Dlaczego warto korzystać z wtórnej analizy danych?

Generalnie wyróżnia się trzy powody – natury ogólnej, metodologicznej i ekonomicznej – dla których warto korzystać z wtórnej analizy danych. Z ogólnego punktu widzenia dzięki danym zebranych w różnym czasie możemy lepiej rozumieć kontekst historyczny, opisywać i wyjaśniać zmiany. Tego typu dane pozwalają także dokonywać porównań na przykład pomiędzy narodami czy społeczeństwami. Jeżeli dysponujemy danymi rzetelnymi i dokładnymi, wówczas mamy możliwość ich replikacji, a także prowadzenia badań longitudinalnych. Wtórna analiza danych może również poprawić jakość pomiaru poprzez poszerzenie zakresu zmiennych użytych w naszym badaniu. Ostatni niezwykle istotny aspekt to koszt uzyskania takich danych. W wielu przypadkach jest on niemal zerowy lub kilkanaście, a nawet kilkadziesiąt razy mniejszy niż koszt przeprowadzenia pierwotnego badania sondażowego⁴. Ważnym kanałem w procesie gromadzenia danych wtórnych coraz częściej staje się Internet.

Dla marketingowców medium to stanowi doskonałe źródło informacji o strukturze rynku i trendach w makrootoczeniu, a także o zachowaniach mikrootoczenia, klientów czy konkurencji⁵. Naukowcom, badaczom, studentom, uczniom i innym zainteresowanym Internet ułatwia w wielu przypadkach bezpłatny dostęp do raportów badawczych. Niemal wszystkie instytuty badawcze mają w Internecie swoje witryny, m.in. CBOS (www.cbos.com.pl), TNS OBOP (www.tns-global.pl), PENTOR (www.pentor.pl), SMG/KRC (www.millwardbrown.com/sites/poland). Wiodące agencje badawcze umożliwiają subskrybentom bieżący dostęp do wielu aktualnych raportów, informując zainteresowanych o opublikowanych właśnie wynikach analiz. Wyszukiwarki na stronach instytucji i agencji badawczych pozwalają szybko i efektywnie odszukiwać interesujące nas dane statystyczne według słów kluczowych. Takie informacje znajdziemy na przykład na stronach Głównego Urzędu Statystycznego (www.stat.gov.pl), Ministerstwa Finansów (www.mf.gov.pl) czy Ministerstwa Nauki i Szkolnictwa Wyższego (www.nauka.gov.pl). Wydawane przez GUS roczniki i inne publikacje należą do jednych z najlepszych i najpełniejszych zasobów danych statystycznych dotyczących wielu aspektów życia gospodarczego i społecznego. Za ich pomocą możemy analizować różnice między poszczególnymi regionami kraju czy Europy. Statystyki te w większości przypadków są dostosowane do standardów obowiązujących w UE, stąd możliwe jest dokonywanie analiz porównawczych. Językoznawcy dzięki internetowym forom dyskusyjnym, listom mailingowym i innym narzędziom komunikacyjnym dostępnym w tym medium mogą badać język internautów. Wiele czatów umożliwia zapis prowadzonych rozmów, dzięki czemu można je później analizować. Socjologowie mający dostęp do serwisów społecznościowych, mogą opisywać zmiany w zakresie tworzenia się i funkcjonowania społeczności, psychologowie – eksplorować blogi jako formy ekspresji i autokreacji, a kulturoznawcy – zgłębiać

⁴ Ch. Frankfort-Nachmias, D. Nachmias, *Metody badawcze w naukach społecznych*, Poznań 2001, s. 322–324.

zmiany w obrębie kultury internetowej. Internet to prawdziwa kopalnia danych, a w dodatku w wielu przypadkach uzyskiwanych bezpłatnie i szybko.

Jak pozyskiwać dane z Internetu?

Najprostszym sposobem na uzyskiwanie danych z Sieci jest korzystanie z wyszukiwarek czy różnego rodzaju baz danych (np. www.polskiebazydanych.pl; www.consumerdata.pl; www.kig.pl). Dane można również czerpać z ulubionych stron polecanych przez innych za pomocą zakładek (bookmarków), jak to ma miejsce w przypadku delicious.com czy stumbleupon.com. Jest to doskonały sposób na szybkie uzyskanie dostępu do danych, w wielu bowiem przypadkach dotarcie do interesujących nas informacji w Internecie to proces żmudny, czasem zakończony niską trafnością uzyskanych wyników, a nawet kompletnym niepowodzeniem. Wyniki „pokazywane” przez wyszukiwarkę na najwyższych miejscach nierzadko nijak się mają do naszych preferencji i oczekiwań. Obecnie główną zasadą klasyfikacji zawartości tekstowej strony WWW jest ranking występowania wyrazów bądź łańcuchów znaków. Odbывается to poprzez indeksowanie w pierwszej kolejności nagłówków strony i podnagłówków – aż po tekstową zawartość strony. Oprócz tego wiodąca wyszukiwarka stosuje tzw. *PageRank*, czyli metodę zliczania odsyłaczy (linków) do danej strony. Wartość punktów odsyłacza prowadzącego do danej strony jest tym większa, im wyższe miejsce w rankingu zajmuje strona z odsyłaczem. Jednakże powyższe metody mają swoje słabe punkty. Wystarczy, że projektant umieści linki na kilkunastu różnych stronach (różnych firm), wzajemnie prowadzące do siebie, aby *PageRank* wzrósł. Poza tym stosuje się zmianę zawartości strony w zależności od oglądającego (inaczej strona wygląda dla użytkownika, inaczej dla wyszukiwarki)⁶.

Menedżerowie, ale także zwykli konsumenci doskonale wiedzą, jak istotnym źródłem danych o przedsiębiorstwach jest Internet. Zgodnie z zasadą: jeśli nie istniejesz w Internecie – nie istniejesz w ogóle, wiele firm zdaje sobie sprawę, że ich witryna WWW w wielu przypadkach decyduje o ich sukcesie bądź klęsce. Brak jakichkolwiek informacji o danym przedsiębiorstwie w Internecie może być dla nas również istotnym sygnałem. Sprawdzając wiarygodność danej firmy, warto sięgnąć do for dyskusyjnych, gdzie nierzadko można znaleźć informacje na temat niecných praktyk prowadzonych w danej firmie czy o nieuczciwym pracodawcy, który na przykład nie wypłaca wynagrodzenia. W wielu przypadkach takie informacje okazują się bezcenne. Fora i listy dyskusyjne są ważnym źródłem danych o charakterze marketingowym. Konsumenci chwalą kupowane przez siebie produkty, ale również jasno i dobitnie wyrażają dezaprobatę, gdy produkt nie spełnił ich oczekiwań. Bagatelizowanie takich informacji byłoby lekkomyślne. Jak słusznie zauważa Magdalena Sobocińska, internetowe blogi pozwalają poznać postawy, hierarchie wartości, gusta, język i sposoby spędzania wolnego czasu przez ludzi młodych⁷. Oczywiście nie tylko młodych, choć

⁵ M. Sobocińska, *Nowe technologie gromadzenia informacji*, [w:] *Badania marketingowe...*, s. 92.

⁶ M. Szpunar, *Internet – medium informacji versus dezinformacji*, „E-Mentor” 2007, nr 2, s. 46.

z pewnością przez tę kategorię wiekową to medium jest zdominowane. Coraz częściej wiele internetowych nisz „zasiedlają” również seniorzy, którzy do niedawna stanowili kategorię niemal całkowicie wykluczoną z globalnej Sieci.

Doskonałym źródłem danych w Internecie są również archiwa prasowe, niestety w wielu przypadkach płatne. Takie archiwa prowadzą np. „Gazeta Wyborcza” (szukaj.wyborcza.pl/archiwum), „Rzeczpospolita” (www.rzeczpospolita.pl/szukaj/archiwum.pl) czy „Polityka” (www.polityka.pl/archive).

Jak korzystać z treści dostępnych w Internecie?

W przypadku korzystania z treści dostępnych w Internecie rodzą się wątpliwości, jak z nich korzystać. Czy prowadząc obserwację na czacie, wywiady w serwisie społecznościowym lub też analizę treści na forum dyskusyjnym lub blogu, należy zwrócić się do internautów o wyrażenie zgody na korzystanie z ich wypowiedzi? Czy też w związku z tym, iż informacje te są publicznie udostępniane, przy zachowaniu zasady anonimowości, nie ma konieczności zwracania się do kogokolwiek o wyrażenie zgody? Tradycyjne reguły etyczne wskazują, że badacze, którzy obserwują ludzi w miejscach publicznych, zazwyczaj nie muszą uzyskiwać od nich zgody, a badania przeprowadzane w sytuacjach, w których oczekuje się jakiegoś stopnia prywatności, takiej zgody wymagają⁸. Roger Wimmer i Joseph Dominick uważają, że jeśli prowadzimy tzw. analizę bierną, tj. wykorzystujemy treści zamieszczone na stronach internetowych, blogach czy forach dyskusyjnych, czyli witrynach dostępnych publicznie, oglądanych przez szeroką publiczność, to taki materiał można dowolnie analizować i cytować bez zgody właściciela danej witryny⁹. Wydaje się, że w tym względzie należy stosować takie same standardy etyczne jak w przypadku prowadzenia badań drogą *offline*, choć na jakiś *uzus* przyjdzie nam pewnie jeszcze poczekać. Zakładanie złej woli badaczy, których celem jest nieszanowanie praw i dóbr osoby badanej, jest błędem, choć i takie sytuacje mogą się zdarzać.

Choć trudno w cyberprzestrzeni jednoznacznie określić, co stanowi przestrzeń prywatną, a co publiczną, gdyż granice te są płynne, pomocą w rozstrzygnięciu dylematów etycznych w wypadku korzystania z treści publikowanych w Sieci mogą być poniższe pytania sformułowane przez Wimmera i Dominicka, a dotyczące forum dyskusyjnego¹⁰:

- 1) czy jest to forum dostępne dla wszystkich,
- 2) czy aby umieścić post, wymagana jest rejestracja lub hasło,
- 3) czy na stronie zawarto zastrzeżenie, że nie wolno publikować treści w niej zawartych,
- 4) czy uczestnicy spodziewają się prywatności w odniesieniu do swoich postów?

⁷ M. Sobocińska, *Nowe technologie...*, s. 94.

⁸ R.D. Wimmer, J.R. Dominick, *Mass media. Metody badań*, Kraków 2008, s. 599.

⁹ Tamże, s. 115.

¹⁰ Tamże.

Jeśli analizujemy forum zamknięte, gdzie wymagana jest rejestracja uczestników, gdy dostępu do niego bronią hasła, a administratorzy danej strony wyraźnie zastrzegają, iż wykorzystywanie treści zawartych na forum jest zabronione albo wymaga specjalnej zgody, to niedotrzymywanie tych warunków jest nieetyczne. Możemy domniemywać, że uczestnicy danego forum nie życzą sobie, by ich treści w jakikolwiek sposób wykorzystywano. Jednocześnie, jak słusznie zauważają Wimmer i Dominick, brak wyraźnego zastrzeżenia dotyczącego ochrony prywatności wcale nie musi oznaczać, iż stronę należy uważać za publiczną, dopuszczającą cytowanie i analizowanie¹¹. Zawsze warto uzyskać zgodę administratora czy moderatora na korzystanie z treści tam zawartych, a jeśli to możliwe – wszystkich członków forum (oczywiście, jeśli forum nie jest bardzo liczne). Sytuacja jest o wiele mniej skomplikowana w przypadku wykorzystywania treści zawartych na blogu, gdyż jego właścicielem najczęściej jest jedna osoba.

Ograniczenia danych wtórnych

Pozyskiwanie danych wtórnych z Internetu ma swoje zalety, ale oczywiście nie jest pobawione wad. Pojawiają się tutaj problemy związane z wiarygodnością zdobytych w ten sposób informacji. David Nicholas twierdzi, iż za wiarygodne informacje w Sieci możemy uznać takie, które zostały potwierdzone w kilku źródłach, czyli jeśli kilka serwisów publikuje tę samą informację, możemy postawić hipotezę, że jest ona prawdziwa. Ten mechanizm weryfikacji nie zawsze się jednak sprawdza, gdyż np. blogerzy czy dziennikarze obywatelscy w wielu przypadkach dublują informacje, stając się sami dla siebie ich źródłem, niejako „dusząc się we własnym sosie”¹². Zbigniew Łęski i Zbigniew Wieczorek społeczność blogerów określają mianem blogosfery. Podkreślają przy tym, że może istnieć sytuacja, gdzie blogujący funkcjonują w hermetycznym świecie blogów własnych oraz swoich znajomych i oceniają świat wyłącznie na podstawie wiadomości dostarczanych przez inne blogi, nie czując potrzeby ich weryfikacji¹³.

Innym sposobem weryfikacji informacji pochodzących z Sieci jest zwrócenie uwagi na to, kto je publikuje. Jeżeli są to podmioty cieszące się autorytetem w rzeczywistości materialnej (np. www.gazeta.pl; www.rzeczpospolita.pl), zaufaniem społecznym (np. www.wosp.org.pl) lub będące znanymi osobistościami na przykład w dziedzinie nauki, wówczas pozyskiwane z ich stron dane nie powinny budzić wątpliwości. Podobnie jest z informacjami publikowanymi na internetowych stronach rządu (np. www.mf.gov.pl), parlamentu (np. www.sejm.gov.pl), instytucji edukacyjnych czy wyższych uczelni (np. www.agh.edu.pl; www.uj.edu.pl)¹⁴.

¹¹ Tamże, s. 116.

¹² M. Szpunar, *Dziennikarstwo obywatelskie w dobie Internetu*, [w:] *Rola informatyki w naukach ekonomicznych i społecznych*, pod red. K. Grysy, Kielce 2008, s. 148.

¹³ Z. Łęski, Z. Wieczorek, *Blogosfera czy „blogosfera” – słów kilka o naturalnej sztuczności blogów*, [w:] *Oblicza Internetu. Architektura komunikacyjna Sieci*, pod red. M. Sokołowskiego, Elbląg 2007, s. 288.

¹⁴ Por. D. Nicholas, *Ocena potrzeb informacyjnych w dobie Internetu*, Warszawa 2001, s. 58.

Wiarygodność zgromadzonych przez nas danych mogą określać takie czynniki, jak dostępność, dokładność, aktualność, odpowiedniość (trafność), porównywalność, a także zastosowana metoda w pomiarze wtórnym¹⁵. Dane opublikowane w Sieci po pewnym czasie mogą być niedostępne. Stąd tak istotne jest podawanie daty dostępu do danej strony internetowej. Część raportów, baz danych czy całych witryn internetowych przestaje w ogóle istnieć, część po pewnym czasie bezpłatnego dostępu wprowadza wyłącznie opcję płatną. Internet cechuje wysoki stopień aktualności publikowanych informacji. Inaczej jest w przypadku tradycyjnych mediów drukowanych, gdzie dostęp do danych informacji jest późniejszy m.in. z powodu czasu trwania samego procesu druku, kolportażu itp. Internet jest medium, które „żyje” szybkim tempem. Nierzadko informacje na danej stronie są kilka, a nawet kilkanaście razy aktualizowane w ciągu dnia. Trafność pozyskanych z Internetu danych wtórnych wiąże się z odpowiedzią na pytanie, czy dotyczą one tego, co chcieliśmy analizować, i czy rzeczywiście pozwalają na rozwiązanie problemu badawczego. Dane niepełne, anonimowe, bez daty publikacji, uzyskane pierwotnie za pomocą niewiadomej metody, na nieznaną nam próbę badawczą itp. należy traktować z dużą dozą nieufności. Jeśli nie dysponujemy informacjami pozwalającymi ustalić, jak przeprowadzono badanie, jaka metoda została wybrana do badania, trudno wnioskować o zaletach i ograniczeniach uzyskanych w ten sposób danych. Ponadto warto pamiętać, że dane wtórne w wielu przypadkach mogą się stać przyczynkiem do zainicjowania badań pierwotnych. Stąd tak istotna jest ich jakość i wiarygodność, gdyż mogą zniekształcać i kierować przysze badania na niepożądany tor.

Internet jako źródło danych o użytkownikach

Pisząc o Internecie jako medium pozwalającym zbierać dane o charakterze wtórnym, nie sposób pominąć jego użytkowników. W Sieci można znaleźć wiele informacji dotyczących internautów, ich zwyczajów, prawidłowości pomiędzy korzystaniem z Internetu przez kobiety i mężczyzn, osoby młodsze i starsze. Możliwy jest niemal bieżący audyt widowni (*site cetric*) na poszczególnych stronach WWW. W Polsce taki pomiar stale prowadzi firma Gemius (audyt.gemius.pl). Dzięki temu można poznać dane o liczbie użytkowników danej strony (*real users*), liczbie jej odsłon, średnim czasie spędzanym na niej przez danego internautę czy wreszcie dowiedzieć się, jak zachowania te się zmieniają.

¹⁵ S. Kaczmarczyk, *Badania marketingowe, metody i techniki*, Warszawa 2002, s. 164–170.

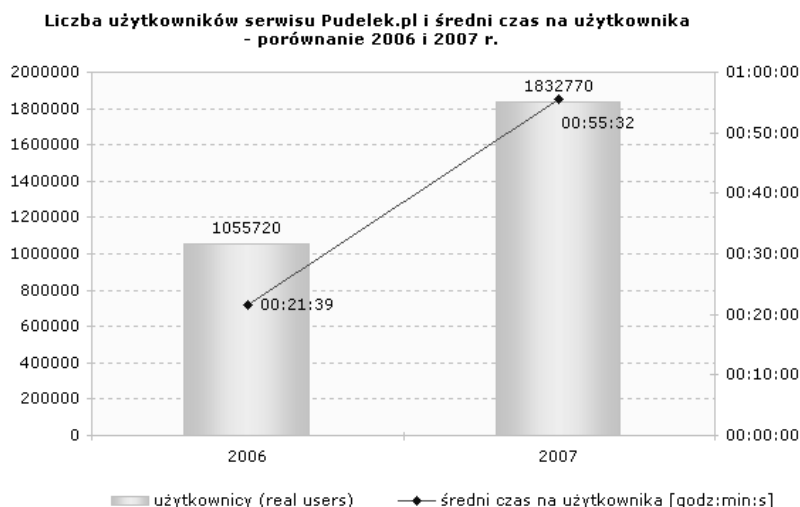
Rysunek 1. Top 10 witryn według zasięgu, dane za sierpień 2008 roku

lp	witryna	użytkownicy (real users)	trend	odśłony	trend
1	Google	13 052 315	▲ 0.7%	2 789 258 408	▲ 0.3%
2	Onet.pl	9 693 896	▼ 0.1%	3 264 124 424	▼ 1%
3	Wirtualna Polska	8 269 051	▼ 0.2%	2 007 623 162	▲ 1.9%
4	Allegro.pl	8 126 152	▲ 0.7%	4 092 991 927	▲ 9.4%
5	nasza-klasa.pl	7 985 445	▲ 0.7%	9 331 267 611	▲ 13.1%
6	INTERIA.PL	7 417 439	▼ 0.7%	1 095 073 971	▼ 1.7%
7	o2.pl	7 176 556	▲ 0.1%	1 003 213 590	▲ 1.1%
8	Gazeta.pl	6 980 606	▼ 0.7%	717 217 730	▲ 4.7%
9	youtube.com	6 977 248	▲ 4.3%	818 077 438	▲ 0.7%
10	wikipedia.org	5 245 329	▼ 2.9%	159 592 166	▼ 25.1%

Źródło: Badanie Megapanel PBI/Gemius, sierpień 2008, grupa: internauci w wieku 7+

Źródło: Badanie Megapanel PBI/Gemius, sierpień 2008, grupa: internauci w wieku 7+.

Warto podkreślić, że Gemius udostępnia te dane w większości przypadków nieodpłatnie. Kopalnią wiedzy o zachowaniach internautów są również publikowane przez Gemiusa raporty dotyczące korzystania z poszczególnych stron WWW, najczęściej zgrupowanych tematycznie, np. witryn o charakterze religijnym, stron o charakterze edukacyjnym czy popularności wśród internautów serwisów plotkarskich.

Rysunek 2. Średni czas korzystania przez użytkownika z serwisu plotkarskiego Pudelek.pl w 2006 i 2007 roku

Źródło: Megapanel PBI/Gemius, czerwiec 2006 i czerwiec 2007.

Jeżeli jesteśmy zainteresowani nie tylko anonimowymi, lecz również konkretnymi użytkownikami na przykład naszej domowej strony internetowej, warto sięgnąć do *Google Analytics*. To niewątpliwie przydatne narzędzie pozwala na ustalenie, w jaki sposób użytkownicy trafili na naszą stronę (bezpośrednio, poprzez wyszukiwarki czy poprzez inne strony), jakich przy tym poszukiwali słów, ile czasu spędzili na naszej witrynie, jaki odsetek odwiedzających stanowili użytkownicy lojalni, tj. tacy, którzy powracali na naszą witrynę, a wreszcie – skąd pochodzą, a tym samym, jakie pole pozostaje nam do zagospodarowania.

Rysunek 3. Przykładowy raport z *Google Analytics*



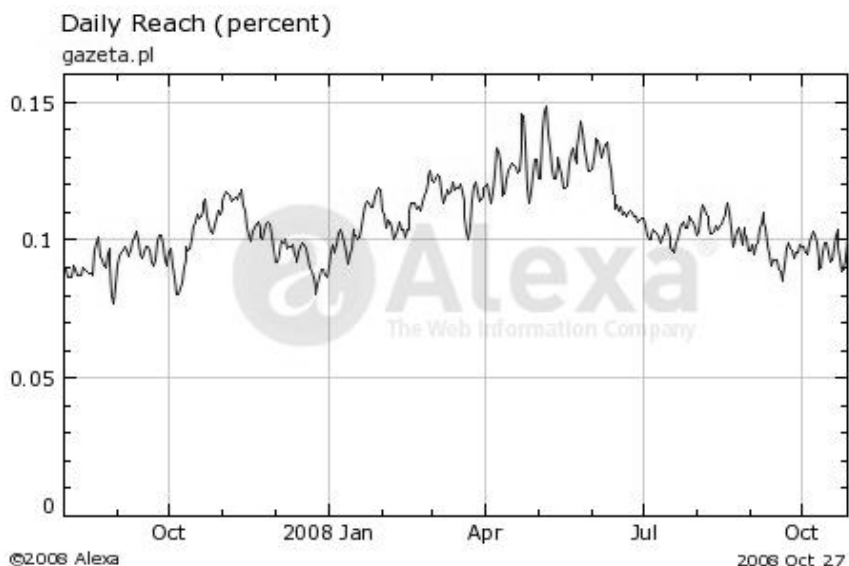
Źródło: www.google.com/analytics/reporting.

Statystyki danej strony internetowej (jej wyszukiwanie poprzez wpisanie adresu internetowego do wyszukiwarki) udostępnia kolejne narzędzie Google – *Google Trends* (www.google.com/trends).

Rysunek 4. Wyszukiwania strony gazeta.pl w okresie styczeń 2004–październik 2008

Źródło: www.google.com/trends.

Pomiarem ruchu na danej witrynie zajmuje się również Alexa (www.alexa.com). Na rysunku 5. zaprezentowano ruch na portalu gazeta.pl w okresie od października 2007 do października 2008. Dzięki Alexie możemy się dowiedzieć, jak wyglądał ruch na danej stronie w przeszłości, jak prezentuje się on obecnie, a jakich tendencji możemy oczekiwać w przyszłości.

Rysunek 5. Ruch internautów na portalu gazeta.pl w okresie październik 2007 – październik 2008

Źródło: www.alexa.com.

Ze stron anglojęzycznych dotyczących badań zbiorowości internautów na szczególną uwagę zasługuje *Pew Internet & American Life Project*, gdzie poszczególne raporty badawcze zgrupowano m.in. w kategoriach: aktywność online, demografia, ewolucja Internetu, korzystanie z mediów i technologii, zdrowie, rodzina, przyjaciele & społeczność, główne wydarzenia, edukacja, praca.

Podsumowanie

Dane zastane – jak niemal każdy inny rodzaj danych – poza niewątpliwymi zaletami mają również swoje ograniczenia. Jednym z nich jest kwestia trafności. Korzystając z tego typu danych, nie mamy gwarancji, że będą one odpowiednie dla naszych celów badawczych¹⁶. Być może osoba, która przed nami je zbierała, miała takie same wątpliwości jak my, jednakże badając tę samą problematykę, można zadawać inne pytania lub inaczej je formułować. Inaczej mówiąc, dane wtórne bardzo często jedynie częściowo pomogą nam w potwierdzeniu lub odrzuceniu naszych hipotez badawczych. Problemy mogą sprawiać również dostęp do tych danych, ich selekcjonowanie i przetwarzanie. Kolejny kłopot wynika z tego, że badacz nie dysponuje wystarczającą wiedzą na temat tego, w jaki sposób dane pierwotne zostały zebrane¹⁷. Informacje o tym, jaką metodą prowadzono badania, na jakiej próbie statystycznej i jaką stosowano technikę badawczą, wiele mówią o ograniczeniach uzyskanych danych. Im osoba pracująca na podstawie danych wtórnych więcej wie na temat sposobu ich pierwotnego zebrania, tym sytuacja jest dla niej korzystniejsza. Brak tego typu informacji utrudnia rzetelną ocenę danych, a w konsekwencji proces ich analizy i opracowywania. Z pewnością warto korzystać z prawdziwej kopalni informacji, jaką jest Internet, jednakże trzeba mieć świadomość ograniczeń uzyskanych tą drogą danych, jak i danych zastanych w ogóle.

¹⁶ E. Babbie, *Badania społeczne...*, s. 304.

¹⁷ Ch. Frankfort-Nachmias, D. Nachmias, *Metody badawcze...*, s. 325.